**Problem Set 4**

**Problem 1.**

Table 5.1:

1. It is a Vertical Listing, rows are attributes, and columns are instances. It is also a Horizontal Listing, columns are attributes, and rows are instances. The unusual structure of this table is each column contains more than one, but three sub-attributes.
2. Queries can be “preprocessing terms tokens nonpositional postings”, or “ effect preprocessing”. Upon the queries, the relations can be extracted by matching header, leftmost column of the table. The false positive can be the tables with effect of other preprocessing except the terms listing on the leftmost column, or effect on totally different attributes with the same preprocessing.

Table 5.3:

1. It is a Vertical Listing, rows are attributes, and columns are instances.
2. Queries can be “computer index compress encoding gaps” or “computer index encoding gaps document ID”. The relations can be extracted by matching the header of column, and leftmost column. The false positive can be the tables referring other words listing encoding gaps and docIDs, since the term “the”, “computer” might not be treated as instances.

Figure 9.8:

1. It is Vertical Listing and Enumeration. It lists several words and their corresponding nearest neighbors.
2. Query can be “word nearest neighbors”. False positive might be the tables listing different instances. The true positive table is buried by massive false positive results.

Figure 12.3:

1. It is Horizontal Listing. Each column of the table represents one model of two unigram language models. It also can be Attribute/Value for each column.
2. Query can be “probability comparison unigram language models”. The query will match the context and header of the table. The false positive will be lots of tables referring intensive words, or several totally different models except the models of Figure 12.3. Adding some other query, like model name or specific word can improve the results.

Table 14.3:

1. It can be Horizontal Listing and can also be Vertical Listing. From the KNN with preprocessing and KNN without preprocessing aspect, it is Horizontal Listing. From the column aspect of either KNN with or without preprocessing part, it is Vertical Listing.
2. Query can be “training testing times KNN classification”. The query can match the context or the contents of the table. The false positive might be tables with other KNN classification rather than preprocessing. Providing specific classification name can avoid such false positive.

Table 16.1:

1. It can be Vertical Listing and Attribute/Value. From Attribute/Value aspect, the subject of the table is application of clustering. From Vertical Listing aspect, the columns are attributes, and rows are instances.
2. Query can be “application clustering information retrieval”. The query matches the leftmost column and context of the table. False positive might be other clustering rather than those mentioned in Table 16.1. Query with specific clustering names can help improve the search results.

Table 16.3(a):

1. It can be Vertical Listing and also be Enumeration. Columns are attributes, row are instances. It lists the document text of documents.
2. Query can be “EM clustering”. Since there is no key words in the table contents contain, only the words in the context of the table can be used as query. False positive might be massive tables related to EM clustering, which are not related the document text specified in the Table 16.3. DocID and document text are useless query words while searching.

**Problem 2.**

Filtering rules: single noun phase and plural noun phases of the form “noun”, “proper noun noun”, “noun noun” and gerund. No Adjective, no proper noun.

Arts ---> music ---> jazz ---> Dixieland jazz

| | |--> bossa nova

| |

| |--> salsa ---> the New York style

| |--> the Puerto Rican bomba

|

|-> theatre ---> lighting ---> lamps

| |--> candles

|

|--> actor ---> a trap door

|--> a person

[arts ---> music](http://www.ieltsbuddy.com/sample-ielts-writing.html)

[arts ---> theatre](http://www.ieltsbuddy.com/essays-on-the-arts.html)

[music ---> jazz](https://en.wikipedia.org/wiki/Music)

[music ---> salsa](http://www.macmillandictionary.com/us/thesaurus-category/american/types-of-music)

[theatre ---> lighting](https://en.wikipedia.org/wiki/Parts_of_a_theatre)

[theatre ---> actor](http://www.razzlebamboom.com/California%20Standards%20-%20Wright%20Brothers%20Musical.pdf)

[jazz ---> Dixieland jazz](https://en.wikipedia.org/wiki/Free_jazz)

[jazz ---> bossa nova](https://en.wikipedia.org/wiki/Jazz)

[salsa ---> the New York Style](https://en.wikipedia.org/wiki/Salsa_(dance))

[salsa ---> the Puerto Rican bomba](https://en.wikipedia.org/wiki/Salsa_music)

[lighting ---> lamps](https://www.reddit.com/r/Fallout/comments/3tuvwq/is_there_a_way_to_craft_non_electrical_lighting/)

[lighting ---> candles](http://americanhistory.si.edu/lighting/19thcent/consq19.htm)

[actor ---> a trap door](http://community.stencyl.com/index.php?topic=4748.5;wap2)

[actor ---> a person](https://books.google.com/books?id=JwQKklQTutIC&pg=PA6&lpg=PA6&dq=%22actor+such+as%22&source=bl&ots=DP9rMjvzfu&sig=XXrHDQsvXdc_FAdZoDAvr4IHkhs&hl=en&sa=X&ved=0ahUKEwjfo_e7n5bMAhXBwj4KHUDMC10Q6AEIPDAJ#v=onepage&q=%22actor%20such%20as%22&f=false)